

UNITED STATES PATENT APPLICATION

for

METHOD FOR THREE-DIMENSIONAL POSITION CALIBRATION OF AUDIO  
SENSORS AND ACTUATORS ON A DISTRIBUTED COMPUTING PLATFORM

Inventors:

VIKAS C. RAYKAR

RAINER W. LIENHART

IGOR V. KOZINTSEV

Prepared by:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN  
12400 Wilshire Boulevard  
Seventh Floor  
Los Angeles, CA 90025-1026  
(408) 720-8300

Attorney Docket No.: 42P17496

## **BACKGROUND OF THE INVENTION**

[0001] Many emerging applications like multi-stream audio/video rendering, hands free voice communication, object localization, and speech enhancement, use multiple sensors and actuators (like multiple microphones/cameras and loudspeakers/displays, respectively). However, much of the current work has focused on setting up all the sensors and actuators on a single platform. Such a setup would require a lot of dedicated hardware. For example, to set up a microphone array on a single general purpose computer, would typically require expensive multichannel sound cards and a central processing unit (CPU) with larger computation power to process all the multiple streams.

[0002] Computing devices such as laptops, personal digital assistants (PDAs), tablets, cellular phones, and camcorders have become pervasive. These devices are equipped with audio-visual sensors (such as microphones and cameras) and actuators (such as loudspeakers and displays). The audio/video sensors on different devices can be used to form a distributed network of sensors. Such an ad-hoc network can be used to capture different audio-visual scenes (events such as business meetings, weddings, or public events) in a distributed fashion and then use all the multiple audio-visual streams for emerging applications. For example, one could imagine using the distributed microphone array formed by laptops of participants during a meeting in place of expensive stand alone speakerphones. Such a network of sensors can also be used to detect, identify, locate and track stationary or moving sources and objects.

[0003] To implement a distributed audio-visual I/O platform, includes placing the sensors, actuators and platforms into a space coordinate system, which includes determining the three-dimensional positions of the sensors and actuators.

### **BRIEF DESCRIPTION OF DRAWINGS**

[0004] The present invention is illustrated by way of example and is not limited by the figures of the accompanying drawings, in which like references indicate similar elements, and in which:

[0005] **Figure 1** illustrates a schematic representation of a distributed computing platform consisting of a group of computing devices.

[0006] **Figure 2** is a flow diagram describing, in greater detail, the process of generating the three-dimensional position calibration of audio sensors and actuators in a distributed computing platform, according to one embodiment of the present invention.

[0007] **Figure 3** illustrates the actuator and sensor clustering process in one embodiment of the present invention.

[0008] **Figure 4** is an example of a chronological time schematic that isolates  $T_s$  and  $T_m$  in one embodiment of the present invention.

[0009] **Figure 5** shows a computing device node which has information regarding the acoustic signal's time of flight (TOF) with respect to multiple nodes in one embodiment of the present invention.

[0010] **Figure 6** illustrates the application of the non-linear least squares (NLS) reliability information to the final calculated node coordinates in one embodiment of the present invention.

## **DETAILED DESCRIPTION**

[0011] Embodiments of a three-dimensional position calibration of audio sensors and actuators in a distributed computing platform are disclosed. In the following description, numerous specific details are set forth. However, it is understood that embodiments may be practiced without these specific details. In other instances, well-known circuits, structures and techniques have not been shown in detail in order not to obscure the understanding of this description.

[0012] Reference throughout this specification to “one embodiment” or “an embodiment” indicate that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment. Thus, the appearances of the phrases “in one embodiment” or “in an embodiment” in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments.

[0013] **Figure 1** illustrates a schematic representation of a distributed computing platform consisting of a group of computing devices (100, 102, 104, 106, and 108). The computing devices include a personal computer (PC), laptop, personal digital assistant (PDA), tablet PC, or other computing devices. In one embodiment, each computing device is equipped with audio actuators 110 (E.g speakers) and audio sensors 112 (E.g. microphones). The audio sensors and actuators are utilized to estimate their respective

physical locations. In one embodiment these locations can be only calculated as relative to each other. In another embodiment these locations can be with reference to one particular computing device said to be located at the origin of a three-dimensional coordinate system. In one embodiment the computing devices can also be equipped with wired or wireless network communication capabilities to communicate with each other.

[0014] Additionally, certain parts of calculations necessary to determine the physical locations of these computing devices can be performed on each individual computing device or performed on a central computing device in different embodiments of the present invention. The central computing device utilized to perform all of the location calculations may be one of the computing devices in the aforementioned group of computing devices in one embodiment. Otherwise, the central computing device is only used for calculations in another embodiment and is not one of the computing devices utilizing actuators and sensors for location calculations.

[0015] For example, in one embodiment, given a set of  $M$  acoustic sensors and  $S$  acoustic actuators in unknown locations, one embodiment estimates their respective three dimensional coordinates. The acoustic actuators are excited using a predetermined calibration signal such as a maximum length sequence or chirp signal, and the time of flight (TOF) of the acoustic signal from emission from the actuator to reception at the sensor is estimated for each pair of the acoustic actuators and sensors. In one embodiment, the TOF for a given pair of actuators and sensors is defined as the time for the acoustic signal to travel from the actuator to the sensor. Measuring the TOF and

knowing the speed of sound in the acoustical medium, the distance between each acoustical signal source and the acoustical sensors can be calculated, thereby determining the three dimensional positions of the actuators and the sensors. This only gives a rough estimate of the actual positions of the actuators and sensors due to systemic and statistical errors inherent within each measurement.

[0016] **Figure 2** is a flow diagram describing, in greater detail, the process of generating the three-dimensional position calibration of audio sensors and actuators in a distributed computing platform, according to one embodiment of the present invention. The flow diagram has a number of steps that are designed to minimize the errors associated with the systemic and statistical errors produced when completing the initial TOF measurements. The process described in the flow diagram of **Figure 2** periodically references the computing devices of the distributed computer platform illustrated in **Figure 1** and refers to each computing device as a node.

[0017] Upon starting **200** the process each actuator attached to each computing device node emits an acoustic signal. These signals can be spaced chronologically in one embodiment of the invention. In another embodiment of the invention multiple actuators can emit acoustic signals simultaneously each signal consisting of a unique frequency or unique pattern. In one embodiment, the acoustic signal may be a maximum length sequence or chirp signal, or another predetermined signal. In one embodiment the group of computing device nodes are given a global timestamp from one of the nodes or from a central computing device to synchronize their time and allow accurate TOF

measurements between all actuators and all sensors. Then for each node, the TOF is measured between that node and all other nodes (202).

[0018] In block 204, the actuator and sensor for each node are clustered together and regarded to be in the same locations. Thus the measured distance (  $\text{TOFs} / (\text{speed of sound})$  ) between two nodes is estimated from the TOF of the actuator of a first node and the sensor of a second node and the TOF of the actuator of the second node and the sensor of the first node. In one embodiment this estimate is the average of the two TOFs. At this point each node is measured as one individual physical location with no distance between the actuator and sensor for each given node. This clustering introduces a limited amount of error into the exact locations of the actuators and sensors but that error is eventually compensated for to achieve precise locations. **Figure 3** illustrates the actuator and sensor clustering process in one embodiment of the present invention. Computing device 300 has an actuator 302 and a sensor 304 located on it. These two devices are clustered 306 with relationship to each other and a central location 308 is calculated to allow for one universal physical location of the actuator 302 and sensor 304 on computing device 300. Additionally, computing device 310 shows another possibility with the actuator 312 and sensor 314 in different locations upon the computing device. Once again the two devices are clustered 316 and a central location 318 is calculated to represent computing device 310. As stated, the discrepancies between the actual physical locations of the actuator and sensor do not pose an issue because adjustments are made to minimize or possibly eliminate these minimal location errors.

[0019] In block 206 of **Figure 2**, a set of linear equations is solved that allows the systemic errors to be estimated from each currently measured TOF to get a more accurate estimation of the TOF between each pair of nodes. The systemic errors that are inherently in each currently measured TOF include the latency associated with actuator emission and the latency associated with capture reception. Computing devices and their actuator and sensor peripherals are fast when executing commands, but not instantaneous. Analog-to-digital and digital-to-analog converters of actuators and sensors of the different nodes are typically unsynchronized. There is a time delay between the time the play/emission command is issued to the actuator and the actual time the emission of the acoustic signal begins (referred to as  $T_s$ ). Furthermore, there also exists a time delay between the time the capture command is issued to the sensor and the actual time the capture/reception of the acoustic signal begins (referred to as  $T_m$ ).  $T_s$  and  $T_m$  can actually vary in time depending on the sound card and processor load of the respective computing device node. These two systemic errors ( $T_s$  and  $T_m$ ) along with the modified TOF using the clustered positions are solved for using a set of linear equations. **Figure 4** is an example of a chronological time schematic that isolates  $T_s$  and  $T_m$  in one embodiment of the present invention. At time 400, the play command is issued. In an embodiment of the invention where all nodes can communicate with each other and have synchronized time stamps the play command will also trigger a capture command at the same instant on a second node. The second node must know when to attempt to capture the signal in order to effectively measure the TOF. At time 402, the capture is started on the second node so  $T_m$  is equal to time 402 minus time 400. At time 404, the emission is started so  $T_s$  is equal to time 404 minus time 400. At time 406, the acoustic signal is



finally captured by the second node, which shows that the true TOF, the time the signal needed to travel through the air to get from the actuator to the sensor is time **406** minus time **404**. Without compensating for the systemic errors  $T_m$  and  $T_s$  each node will have a false assumption as to the true TOF.

[0020] Due to uncertainty in operating conditions of the system as well as external factors it is not uncommon to have certain nodes with incomplete sets of data. In other words, one node might not have the entire set of TOFs for all other nodes. In the case of missing and incomplete data for a node there exists a method to create the rest of the TOFs and subsequent pair-wise node distances. In block **208** of **Figure 2**, the missing data points for a given node can be estimated based on current data received through trilateration. As long as a given node with missing information to node X has at least information relating to four other nodes with TOFs to node X in a two-dimensional environment or five other nodes in a three-dimensional environment, an estimate of the TOF of the nodes with missing information can be calculated. **Figure 5** shows a computing device node A which has information regarding the acoustic signal's TOF with respect to nodes B, C, E, F, G, H, and I in one embodiment of the present invention. It is missing information from node D. Considering this to be a three-dimensional scenario, if at least a set of five of the known nodes out of the set of nodes B, C, E, F, G, H, and I have information regarding node D, then using trilateration node A can obtain the information relating to node D.

[0021] Once the matrix of pair-wise node TOFs is complete or filled in with as much information as possible the next step in one embodiment of the present invention is to calculate the estimated physical position of every node with multidimensional scaling (MDS) using the set of pair-wise node TOFs in block **210** of **Figure 2**. MDS will give estimated coordinates of the clustered center of each node's actuator-sensor pair. In one embodiment one node is set to the origin of the three-dimensional coordinate system and all other nodes are given coordinates relative to the origin. The MDS approach may be used to determine the coordinates from, in one embodiment, the Euclidean distance matrix. The approach involves converting the symmetric pair-wise distance matrix to a matrix of scalar products with respect to some origin and then performing a singular value decomposition to obtain the matrix of coordinates. The matrix coordinates in turn, may be used as the initial guess or estimate of the coordinates for the respective computing device nodes, and the clustered location of the actuator and sensor located on them.

[0022] In block **212** of **Figure 2** a TOF-based nonlinear least squares (NLS) computation is used to determine the individual coordinates of the actuator and sensor of each node. In one embodiment, the TOF-based NLS computation considers the TOFs measured in block **202**, the MDS coordinate results from block **210**, and  $T_m$  and  $T_s$  from block **206**. The NLS computation also reveals a probability assessment that determines the reliability of each node's coordinates using the variance.

[0023] In block **214** of **Figure 2** a Time Difference of Flight (TDOF) NLS computation is used to determine the individual coordinates of the actuator and sensor of each node. The TDOF method is unlike the TOF method. In one embodiment a TDOF method uses three nodes per calculation. The first node excites its actuator and an acoustic signal propagates from it. Two separate nodes (the second and third nodes) each receive the acoustic signal from the first node a short time later. In this scenario there are two recorded TOFs, the TOF between the first node and the second node and the TOF between the first node and the third node. The TDOF is the difference in time between the two TOFs. This is a more indirect way of estimating the coordinate system but in many ways more accurate under certain conditions because the difference in reception times only needs to take into account one of the systemic errors, the sensor error  $T_m$ . Thus, reducing the number of variables allows for a different but possibly more accurate calculation of node coordinates using TDOF. Therefore, in one embodiment, the TDOF-based NLS computation considers the TDOFs calculated from all TOF measurements in block **202**, the MDS coordinate result from block **210**, and  $T_m$  from block **206**. Once again, the NLS computation also reveals a probability assessment that determines the reliability of each node's coordinates using the variance.

[0024] Finally, in block **216** of **Figure 2**, the final coordinates of each individual actuator and sensor on each node are calculated using the coordinate position information and reliability information obtained from the TOF-based NLS computation in block **212** and the TDOF-based NLS computation in block **214** and the process is finished **218**. **Figure 6** illustrates the application of the NLS reliability information to the final calculated node

coordinates in one embodiment of the present invention. In this example point A is the calculated coordinates obtained from the TOF-based NLS computation and ellipse 600 is the variance that shows the reliability of the TOF-based estimate. Point B is the calculated coordinates obtained from the TDOF-based NLS computation and ellipse 602 is the variance that shows the reliability of the TDOF-based estimate. When combining the coordinates together taking into account the reliability of each set the final calculated physical location ends up as coordinate C. Combining both the TOF-based method with the TDOF-based method creates a more accurate estimated end result.

[0025] The techniques described above can be stored in the memory of one of the computing devices as a set of instructions to be executed. In addition, the instructions to perform the processes described above could alternatively be stored on other forms of computer and/or machine-readable media, including magnetic and optical disks. Further, the instructions can be downloaded into a computing device over a data network in a form of compiled and linked version.

[0026] Alternatively, the logic to perform the techniques as discussed above, could be implemented in additional computer and/or machine readable media, such as discrete hardware components as large-scale integrated circuits (LSI's), application-specific integrated circuits (ASIC's), firmware such as electrically erasable programmable read-only memory (EEPROM's); and electrical, optical, acoustical and other forms of propagated signals (e.g., carrier waves, infrared signals, digital signals, etc.); etc.

[0027] These embodiments have been described with reference to specific exemplary embodiments thereof. It will, however, be evident to persons having the benefit of this disclosure that various modifications and changes may be made to these embodiments without departing from the broader spirit and scope of the embodiments described herein. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.